International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2017)
R. Lindeman, G. Bruder, and D. Iwai (Editors)

Evaluating the Effects of Hand-gesture-based Interaction with Virtual Content in a 360° Movie

Humayun Khan[†1], Gun Lee[2], Simon Hoermann[1], Rory M.S. Clifford[1], Mark Billinghurst[2], Robert W. Lindeman[1]

[1]Human Interface Technology Laboratory New Zealand, University of Canterbury, New Zealand
[2]Empathic Computing Laboratory, School of ITMS, University of South Australia, Australia

Abstract

Head-mounted displays are becoming increasingly popular as home entertainment devices for viewing 360° movies. This paper explores the effects of adding gesture interaction with virtual content and two different hand-visualisation modes for 360° movie watching experience. The system in the study comprises of a Leap Motion sensor to track the user's hand and finger motions, in combination with a SoftKinetic RGB-D camera to capture the texture of the hands and arms. A 360° panoramic movie with embedded virtual objects was used as content. Four conditions, displaying either a point-cloud of the real hand or a rigged computer-generated hand, with and without interaction, were evaluated. Presence, agency, embodiment, and ownership, as well as the overall participant preference were measured. Results showed that participants had a strong preference for the conditions with interactive virtual content, and they felt stronger embodiment and ownership. The comparison of the two hand visualisations showed that the display of the real hand elicited stronger ownership. There was no overall difference for presence between the four conditions. These findings suggest that adding interaction with virtual content could be beneficial to the overall user experience, and that interaction should be performed using the real hand visualisation instead of the virtual hand if higher ownership is desired.

Categories and Subject Descriptors (according to ACM CCS): Information Interfaces and Presentation [H.5.1]: Multimedia Information Systems—Artificial, Augmented, and Virtual Realities; Computer Graphics [I.3.3]: Three-Dimensional Graphics and Realism—Virtual Reality

1. Introduction

Low cost Head-Mounted Displays (HMD) available in the consumer market present an opportunity for movie makers to tell their stories in a novel medium. Conventional movie watching at home on a TV, laptop, or projector generally occupies 60° or less of the viewer's Field of View (FoV), limiting the movie experience. The viewer's presence in the movie primarily depends on the narrative of the story, as well as the size of the display and audio configuration. In 360° movies using an HMD, the viewer is fully immersed in the digital content. The ability to interact with the digital content could further affect the user's presence and sense of embodiment within the virtual environment.

Among the various different Virtual Reality (VR) rendering techniques, 360° panoramic video is often used for movie production, as the recording hardware is easily available to capture the real-world in a 360° video. However, the captured video lacks user interaction. This can be added by embedding interactive 3D objects within the 360° video, to enhance the presence and the user experience within the system. The added 3D objects should be relevant to the 360° movie, so as not to detract from the movie narrative. There are different amounts of interactivity that can be added to a 360° video. Dolan and Parets [DP17] defined four different quadrants of interactive Influence and Existence in the 360° video format. A user's Influence in the video could be Active or Passive. In the Active case, the user effects the narrative of the story, whereas in the Passive scenario, the user has no influence on the outcome of the story. Existence also has two cases: Observant or Participant. In an Observant case, the user remains outside of the movie and does not interact with the digital content of the movie. However, in the Participant case, the user exist in the movie and is able to interact with the digital content. The four combinations of Influence and Existence therefore are (1) Observant Passive, (2) Observant Active, (3) Participant Active, and (4) Participant Passive. For this paper, we performed an experiment using Participant Passive, immersing the viewer in a 360° video, providing interactivity with digital content, but has no influence on the narrative.

† humayun.khan@pg.canterbury.ac.nz

Interactivity in a 360° movie can be added with tangible user interface devices such as the HTC Vive controller or Oculus Touch, or alternatively, can be included using a Bare-Hand Interaction (BHI) technique. Since this research is focused on improving home entertainment system, the BHI approach was selected. To support BHI, a Leap Motion controller was used to accurately track the user's hands and allow interaction with the embedded digital content. Real hand information is included by combining a RGB-D SoftKinetic camera with the Leap Motion controller. The SoftKinetic camera captured hand texture as well as depth information in real-time. The Leap Motion controller and SoftKinetic camera were calibrated together, using the tracking information from both cameras. The SoftKinetic hand tracker gave the fingertip positions, which were used to find optimal rotation and translation to the corresponding Leap Motion fingertip positions using Singular Value Decomposition (SVD). This enabled the inclusion of real hand interaction in the designed system.

In this work, a Mixed Reality (MR) system was designed and developed that allows users to interact with the digital content in a 360° VR movie using their bare hands. An experiment was performed using the developed system to examine the effect of hand-gesture-based interaction and hand appearance on presence and embodiment in 360° VR movies. Users played a Participant Passive role in the movie and were able to interact with the movie without changing the narrative.

The main contributions of the paper are:

- A novel method to provide real hand interaction in 360° movies, by combining a Leap Motion controller and a SoftKinetic, DS325 camera. The combination used Leap Motion's hand tracking and SoftKinetic's hand visualisation technologies.
- A user study investigating the effect of interaction and representation of user's body on the senses of presence and embodiment in 360° VR movies.

2. Related Work

There is some existing research which explores using MR for cinematic experiences. Cheok et al. [CWY*02] created an interactive theatre system using an external camera array, consisting of 15 cameras and an HMD. The system captured the user in real-time and rendered them in an MR environment, where they could interact with virtual content. The system supported three modes of interaction. The first mode was an outdoor theatre land exploration mode, which allowed users to walk around in an outdoor environment wearing an HMD. The virtual world was overlaid on the physical world and was seen through the first person perspective in the HMD. The second mode was an Augmented Reality (AR) theatre land exploration mode which embedded a virtual theatre in the physical environment by merging the AR and VR worlds. It also supported interaction with virtual objects. The third mode was a virtual interactive theatre mode which featured a fully immersive cinematic environment. A live actor in the physical world was captured by an external camera system and embedded in the virtual world in real-time. While this MR system was conceptually unique, it required an external camera system (15 cameras) which is not suitable for a home entertainment system.

Tenmoku et al. [TIS*06] and Ichikari et al. [IKT*10] proposed a work flow to insert computer graphics animation data, motion capture data and 3D video data into special movie-making software. The position and movement of a camera in the real world could be incorporated into a 3D model, representing the physical location of the filming. Using the 3D model of the scenery and the animation data, the movie director was able to plan ahead the movements and positions of the camera for the shot to be filmed. These systems could be used to create an offline MR cinematic environment, but would not be useful for a real-time cinematic environment.

Research conducted by Lok et al. [LNWB03] studied the effects of self-avatar fidelity on task performance and the user's sense of presence. The study had three different self-avatar representations. The Slater-Usoh-Steed (SUS) questionnaire was used to measure presence [UCAS00]. The quantitative data showed no significant difference for self-avatar fidelity. However, the qualitative feedback indicated that the users felt a higher sense of presence with higher fidelity self-avatars. In their discussion, Lok et al. suggest that there might be a difference in the sense of presence with the higher fidelity self-avatar. In other work, Usoh et al. [UAW*99] hypothesised, with regard to the visual fidelity of an avatar, that "substantial potential presence gains can be had from tracking all limbs and customising avatar appearance".

Bruder et. al [BSRH09] blended the user's actual hands into VR and studied its effects on the sense of presence. The user's hands were segmented using chroma-keying with an egocentric camera and displayed in an HMD in real-time. Their experiment compared two conditions: with or without hand visualisation. The study used the SUS questionnaire to measure presence [UCAS00]. Results indicated that users felt a higher sense of presence on seeing their hands in the virtual environment.

Work by McGill et al. [MBMSB15] studied the effects of interaction with real objects and peripherals on a user's sense of presence while using an HMD. The chroma-keying method was used to segment and blend the hands and periphery into VR. This study compared interaction in the selective reality (partial or minimal blending) with interaction in full-view reality. The Igroup Presence Questionnaire (IPQ) was used for measuring the user's sense of presence [SFR01]. The results indicated a higher sense of presence for partial reality than the full-view reality.

These works show that there is good potential for increase in presence by including interaction and higher fidelity (more realistic) self-avatars.

Research by Argelaguet et al. [AHTL16] studied the sense of agency and ownership of having interaction in a virtual environment, using three different hand model representations (realistic, iconic, and abstract hands). They found that the sense of agency is related to the virtual hand control and the efficiency of performing a task in a virtual environment. They also found that the sense of ownership is mainly related to the visual appearance of the virtual hand. Results of the study showed that the sense of agency was stronger for abstract and iconic hand models than realistic hand models. In contrast, the sense of ownership was stronger for more realistic hands than iconic and abstract hands. The results differed from Lin et al. [LJ16] study who did not find any effect of hand appearance on the sense of agency. In Lin et al.'s study, users were shown six different hand models: realistic hand, toony hand, very toony hand, zombie hand, robot hand and wooden block. The sense of agency remained intact even for non-realistic avatars. However, more realistic hands showed a stronger sense of ownership. For the sense of ownership, results from both studies were similar.

While previous work has been done on interaction with movies, the Participant Passive scenario has not been explored in depth with a 360° movie watching experience. In the following sections, we present a system based on the Participant Passive scenario to explore this domain.

3. System

A prototype was developed to produce an immersive cinematic environment which allows hand-based interaction with 3D objects. Interaction with real textured hands was built by combining the output from two hand trackers, the Leap Motion controller and the SoftKinetic RGB-D camera. The Leap Motion provided more accurate hand tracking, while the SoftKinetic provided higher visual fidelity. For the two systems to complement each other, the output from both were combined in Unity3D. In this configuration, the user's hands are precisely tracked in real-time simultaneously with the skin texture of the user. An immersive VR environment was created in Unity3D using a sphere primitive and a 360° video player (Easy Movie texture plugin [Jae17]). Finally, 3D objects rendered in real-time were embedded within the 360° video to provide an interactive user experience.

3.1. Hardware

In addition to the Leap Motion and SoftKinetic camera, the prototype included an Oculus Rift DK2 HMD and a PC. VR content was displayed on HMD. The PC used to run the system had an Intel i7-6700, 3.4GHz processor, 16GB RAM, and an NVIDIA GTX980 graphics card. To attach the hand tracking sensors onto the HMD, we 3D printed a SoftKinetic bracket [Rea17]. The attachment setup for the SoftKinetic camera is shown in Fig. 1a. Similarly, a Leap Motion controller mount [Lea17] was 3D printed and affixed to the top of the SoftKinetic bracket (Fig. 1b), and the resulting hardware setup is shown in Fig. 1c.

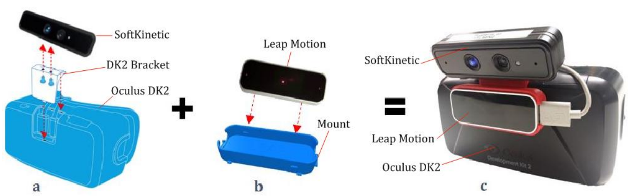


Figure 1: *(a) SoftKinetic, DS325 and attachment bracket to Oculus Rift DK2 (b) Leap Motion controller and mount (c) Combined system hardware.*

3.2. Software

3.2.1. Combining the hand trackers

With the two sensors mounted on the HMD, the alignment between the two sensors is calibrated in software by capturing the hand tracking data from each camera using the SoftKinetic Depth-Sense SDK and Leap Motion Unity3D asset. The captured data from the two cameras was read into the Unity3D game engine to be processed and visualised. To align the two sensors, the transformation matrix between the two data sets was computed based on a corresponding point matching algorithm based on singular value decomposition (SVD) [BM92]. The computed transformation was then applied to the SoftKinetic's data to align it with the Leap Motion's tracking data. With the two sensors aligned, the texture of the real hand from the SoftKinetic's RGB-D camera was then used for visualising the hand, and the Leap Motion's physics-based hand was used for interaction with the virtual content. Since the relative position between the two cameras remained fixed, the transform was only computed once prior to the use of the system.

The corresponding point matching algorithm required at least three corresponding points from each camera. We used Bai et al.'s [BGESB13] algorithm to track six feature points of a user's hand from the SoftKinetic camera, five finger tips and one palm position. The Leap Motion hand tracking results also provided the positions of the six corresponding points for alignment (Fig. 2). Using this set of six corresponding points between the two tracked data sets allowed us to compute the transformation matrix between them.

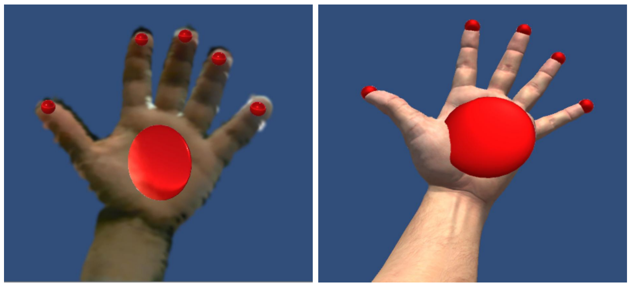


Figure 2: *SoftKinetic (left) and Leap Motion (right) hand tracking data and corresponding points*

The details of the alignment algorithm is explained below step by step, and the whole process is illustrated in the flowchart in Fig. 3.

Alignment Algorithm:

1. First, the centroids (center points) were calculated for both data sets. Six SoftKinetic points were $SKPoint_i$ and six Leap Motion points were $LMPoint_i$, where $N = 6$.

$$SKPoint_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \qquad LMPoint_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}$$

$$centroid_{SK} = \frac{1}{N}\sum_{i=1}^{N} SKPoint_i \qquad centroid_{LM} = \frac{1}{N}\sum_{i=1}^{N} LMPoint_i$$

2. Then, the co-variance matrix, H, was computed.

$$H = \sum_{i=1}^{N} (SKPoint_i - centroid_{SK})(LMPoint_i - centroid_{LM})^T$$

3. The co-variance matrix, H, was factorised using singular value decomposition.

$$[U,S,V] = SVD(H)$$

4. From the factorised H matrix, rotation matrix, R, was computed by performing a dot product between V and transposing of the U matrix which were obtained from SVD.

$$R = VU^T$$

5. The computed R matrix was the rotation matrix if the determinant of R was greater than zero. If it was less than zero, it was a reflection case, and the third column of the R matrix was multiplied by -1 which gives the rotation matrix.

$$\begin{aligned} &\quad if\ determinant(R)\ <\ 0 \\ &\qquad multiply\ 3rd\ column\ of\ R\ by\ -1 \\ &else \\ &\qquad R\ remains\ the\ same \end{aligned}$$

6. The translation matrix, t, was computed using the following equation:

$$t = -R * centroid_{SK} + centroid_{LM}$$

After obtaining the rotation and the translation matrices, they were applied on the initial six points to compute the Root Mean Square Error value, RMSE, which helped to evaluate the accuracy of calculated rotation and translation matrices. This was calculated using the following equation:

$$RMSE = \sum_{i=1}^{N} ||R * SKPoint_i + t - LMPoint_i||^2$$

If the RMSE value was less than the threshold value, the optimal solution was found. If the value was greater than the threshold value, the process was repeated from steps 1 to 6 with different SoftKinetic and Leap Motion data sets until the optimal solution was found.

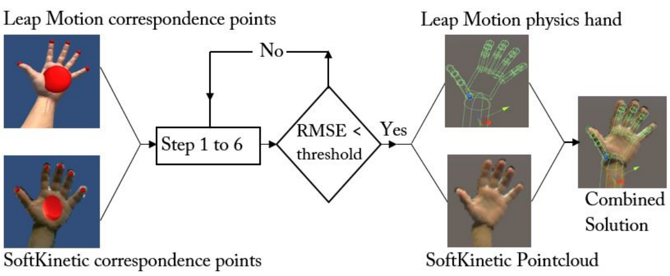


Figure 3: *Flow chart illustrating point cloud alignment*

3.2.2. 360° Spherical Video Player

The 360° video based immersive VR environment was created by combining Unity3D's sphere primitive with a video player. Most of the 360° videos available online use equirectangular mapping which is primarily a spherical coordinate system projected onto a planar coordinate system. 360° spherical video playback in Unity3D is achieved by texture mapping the equirectangular planar videos onto a sphere primitive. The Easy Movie Texture plugin [Jae17] was used as a video player asset in Unity3D. User immersion is achieved by positioning the virtual camera at the centre of the video sphere.

3.2.3. Hand Gestures and 3D Interactive Objects

To create an interactive experience with the system, interactive 3D objects need to be added. As previously mentioned, the objects should have some relevance to the 360° movie being played to improve the user experience. Another constraint that needs to be considered is the space where these objects can be added. The 3D objects should be added in the space between the viewer and the spherical movie, so they add to the movie experience and not disrupt presence.

The 360° video selected for our prototype immersive VR system was a calm hot air balloon ride with ambient music. The video was shot from just outside the hot air balloon, so in-air interaction was more plausible. Through brainstorming, we ideated a convincing in-air interaction which was feeding birds flying around the balloon. To add flying birds, a rigged 3D bird model [WDa17] was used which could perform basic behaviours such as flying, gliding and diving. With the 3D bird models spawned inside the spherical video, basic flying, gliding and diving actions were created within the interaction space. A feeding interaction using a hand gesture was added to the scene. The hand gesture was to open and close the downward facing hand, as shown in Fig. 4. With this gesture, food would appear from the hand position, falling downward towards the ground. The birds were programmed to react to feeding by diving towards the food to catch it. The full feeding interaction is shown in Fig. 5.

(a) *Closed hand facing downwards* **(b)** *Opened hand creating food*

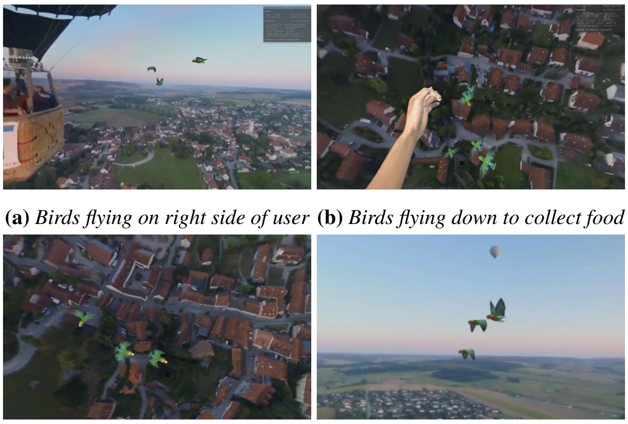
Figure 4: *Hand gesture to create food*

4. User Study

Using the designed immersive 360° movie VR environment, a user experiment was conducted to investigate the effect of interaction and hand appearance on user presence and embodiment in a 360° VR movie.

4.1. Goals and Hypotheses

The first goal of the experiment was to find out whether interaction, compared to no-interaction, increases users' presence and embodiment in the movie. The second goal was to compare the virtual



(a) *Birds flying on right side of user* **(b)** *Birds flying down to collect food*

(c) *Birds flying back with food* **(d)** *Restart normal flying pattern*

Figure 5: *Birds and the feeding interaction*

hands to real hands to find which one creates a higher sense of presence and embodiment. The hypotheses for the experiment are as follows:
[H1] User interaction with the immersive VR movie produces a higher sense of presence and embodiment than without interaction.
[H2] Seeing real hands for interaction produces a higher sense of presence and embodiment than seeing virtual CG (Computer Graphics) hands.

4.2. Setup

The experiment was carried out in a dedicated room as shown in Fig. 6 below.

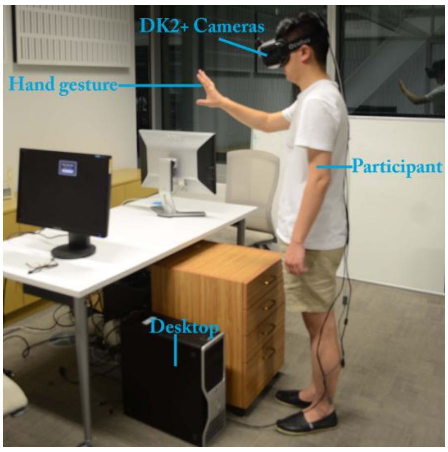


Figure 6: *Experimental setup*

The 360° VR video shown was two minutes long, giving participants sample time to observe the surroundings and to perform interaction within the virtual environment. For the virtual CG hands, users were shown six different hand models to choose from (Fig. 7). Participants were asked to select the model that closely resembled their own hands.

Figure 7: *The six virtual CG Hand Models. Female hands are on the left and male hands are on the right.*

4.3. Participants

There were 32 participants (16 female), all undergraduate and graduate students, recruited for the experiment. Their ages ranged from 19 to 42 years (mean: 26.97, SD: 6.32). Ten of the participants had never used an HMD before, 11 had used it a few times a year, and 11 were more frequent users (daily, weekly, or monthly). All participants gave written consent to the experiment and were reimbursed for their time with a $ 10 voucher.

4.4. Experimental Design

The experiment was in a within-subjects, 2x2 factorial design and using a 4x4 balanced Latin square to counter-balance order effects. There were two independent variables (IV) for the experiment, *interaction* and *hand appearance*. Each IV had two levels. For the interaction IV, the levels were *with interaction* and *without interaction*. For the hand appearance IV, the levels were *real hands* and *virtual CG hands*. In total there were four conditions: **Condition A**, participants were shown their real hands in the movie, and they could interact with the VR environment; **Condition B**, participants were shown their real hands in the movie, but they could not interact with the VR environment; **Condition C**, participants were shown the virtual CG hands in the movie, and they could interact with the VR environment; **Condition D**, participants were shown the virtual CG hands in the movie, but they could not interact with the VR environment.

4.5. Measures

For each condition, there were two measures recorded using a questionnaire: Sense of presence and sense of embodiment. In a post-experiment questionnaire, participants were also asked to rank the four conditions based on their preference.

Sense of Presence

The sense of presence was measured using the Igroup Presence Questionnaire (IPQ) [SFR01], which has been widely used in previous research to measure presence [MBMSB15], [KEB*04]. The IPQ is composed of 14 items, which are rated on a seven-point Likert scale. These 14 items are further divided into three sub-scales (spatial presence, realism and involvement) and one general item (general presence). The three sub-scales and the general item are specified below.

- **Spatial Presence:** The sense of being physically present in the virtual environment.
- **Involvement:** Measuring the attention devoted to the virtual environment and the involvement experienced.
- **Experienced Realism:** Measuring the subjective experience of realism in the virtual environment.
- **General Presence:** Assessing the general "sense of being there".

These three scales are independent of each other. The fourth item, general presence, has an effect on all three sub-scales, especially the perceived spatial presence. For the experiment, IPQ was used to measure presence for all four conditions.

Sense of Embodiment

The sense of embodiment was measured using a questionnaire by Argelaguet et al. [AHTL16], which measures embodiment using the two dimensions of sense of agency and sense of ownership. Sense of agency is the feeling of being in control of either the virtual hands or the real hands. Sense of ownership is the feeling that the representation of the hands belongs to one's own body.

4.6. Procedure

The experiment was reviewed and approved by the university's Human Ethics Committee. As participants arrived, they were asked to read an information sheet and consent form, and sign it if they agreed. Next, they completed a pre-experiment questionnaire asking for their age, gender, and any prior experience with HMDs. They were provided with an outline of the experiment and the background, followed by a learning session to become familiar with the interaction technique to be used during the experiment. Each participant went through the four conditions in a sequence specified by the balanced Latin square design. At the end of each condition, the participants answered a post-condition questionnaire, and they rated their presence and embodiment experience. After completing all four conditions, they answered a post-experiment questionnaire to provide feedback on their overall experience. Lastly, there was a debriefing session to clarify any issues in the questionnaire responses. The study duration was about 40 minutes.

4.7. Task

The task involved participants watching the hot air balloon ride video under each condition. Before each trial, they were told whether they could interact with the birds or not. In the *with interaction* case, they were able to feed the birds and for the *without interaction* case, they were able to see their hands but there was no feeding interaction. The birds remained flying for the *with interaction* case, and for *without interaction* conditions, they dived down to catch the food.

5. Results

This section presents the results obtained from the experiment. The data collected from the questionnaire was in ordinal scale, hence non-parametric tests were used to interpret it. With two IVs in the experiment, interaction and hand appearance, there were two main effects and one interaction effect between the IVs. In order to conduct a factorial analysis on ordinal data, an Aligned Rank Transform (ART) was applied to the ordinal data [WFGH11] before performing a two-way repeated measures analysis of variance (ANOVA) for both presence and embodiment questionnaires ($\alpha = 0.05$).

5.1. Presence

For the overall IPQ questionnaire results, no significant main effect of interaction ($F(1,31) = 0.530$, $p = 0.472$) and no significant main effect of hand appearance (real hands or CG hands) ($F(1,31) = 2.221$, $p = 0.146$). There was also no interaction effect between the two IVs ($F(1,31) = 0.027$, $p = 0.870$).

To further analyse the IPQ results, a two-way repeated measure ANOVA was applied on the four IPQ sub-scales of General Presence, Spatial Presence, Involvement, and Experienced Realism, with the results shown in the Table 1. In summary, the interaction factor had a positive significant main effect on the General Presence and Involvement sub-scales, while hand appearance had a significant main effect on the Involvement and Experienced Realism sub-scales.

Measure	Interaction or No Interaction	Real Hands or CG Hands	Interaction * Hand appearance
IPQ, Overall	**$p = 0.472$, $Mean_{Interaction} = 3.87$, $Mean_{NoInteraction} = 3.75$,**	**$p = 0.146$, $Mean_{RealHands} = 3.82$, $Mean_{CGHands} = 3.81$**	**$p = 0.870$**
IPQ, General Presence	$p = 0.016$, $Mean_{Interaction} = 4.87$, $Mean_{NoInteraction} = 4.39$	$p = 0.887$, $Mean_{RealHands} = 4.63$, $Mean_{CGHands} = 4.64$	$p = 0.993$
IPQ, Spatial Presence	$p = 0.330$, $Mean_{Interaction} = 3.75$, $Mean_{NoInteraction} = 3.65$	$p = 0.954$, $Mean_{RealHands} = 3.71$, $Mean_{CGHands} = 3.70$	$p = 0.918$
IPQ, Involvement	$p < 0.001$, $Mean_{Interaction} = 4.51$, $Mean_{NoInteraction} = 4.00$	$p = 0.005$, $Mean_{RealHands} = 4.44$, $Mean_{CGHands} = 4.08$	$p = 0.001$
IPQ, Realism	$p = 0.565$, $Mean_{Interaction} = 3.55$, $Mean_{NoInteraction} = 3.48$	$p = 0.049$, $Mean_{RealHands} = 3.56$, $Mean_{CGHands} = 3.47$	$p = 0.770$

Table 1: *Two-way repeated measures ANOVA on the four IPQ sub-scales, green cells denote $p < 0.05$.*

5.2. Embodiment

Embodiment was measured by evaluating the user's sense of agency and sense of ownership with a questionnaire. The questionnaire was adopted from Argelaguet et al. [AHTL16], and both agency and ownership were measured separately.

5.2.1. Agency

For agency, a significant main effect of interaction (F(1,31) = 12.256, p = 0.001) was found, yet no significant main effect for hand appearance (F(1,31) = 0.522, p = 0.476) was found. There was no interaction effect between the two IVs (F(1,31) = 0.006, p = 0.938). The box-plot for agency results is shown in Fig. 8.

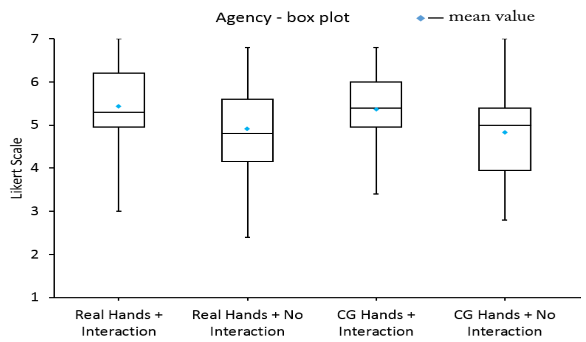


Figure 8: *Agency box-plot*

5.2.2. Ownership

For ownership, we found significant main effects for both factors, interaction (F(1,31) = 9.290, p = 0.005), and hand appearance (F(1,31) = 11.255, p = 0.002). There was no interaction effect between the two IVs (F(1,31) = 0.124, p = 0.727). The box-plot for ownership results is shown in Fig. 9.

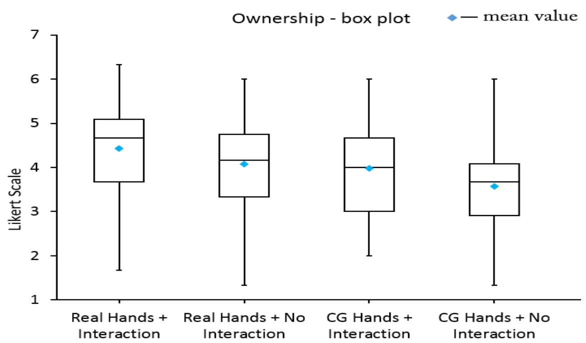


Figure 9: *Ownership box-plot*

5.3. Post-experiment Questionnaire

In a post-experiment questionnaire, the participants were asked which condition they preferred for watching a 360° movie. In the answer, they had to rank the four conditions from 1 (best) to 4 (worst). Friedman test on the ranking data showed a significant difference (χ^2 (3) = 48.0, p < 0.001), so post hoc analysis was performed on each condition pair using Wilcoxon signed-rank tests and a Bonferroni correction applied (significance at p<0.00833). A significant difference was found for four pairs, Condition A and Condition B (Z = -4.580, p < 0.001), Condition A and Condition D (Z = -2.868, p = 0.004), Condition B and Condition C (Z = -4.614, p < 0.001), Condition C and Condition D (Z = -4.580, p < 0.001). The bar chart for the ranking results is shown in Fig. 10. They preferred the CG hands with interaction (Mean Rank: 1.5) closely followed by the real hands with interaction (Mean Rank: 2.0). The display of the CG hands without interaction (Mean Rank: 3.00) and real hands without interaction (Mean Rank: 3.5) were preferred the least.

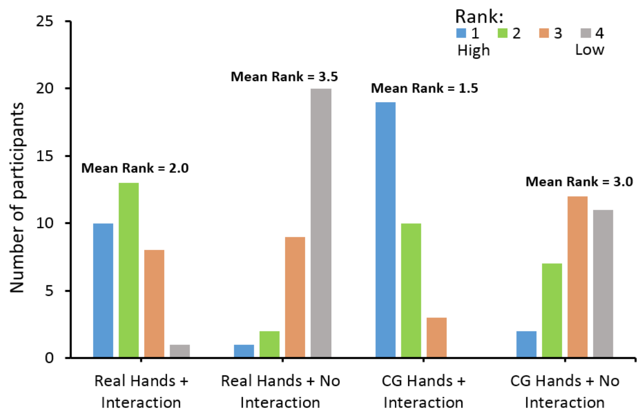


Figure 10: *Conditions ranking bar graph*

6. Discussion

The results show that BHI improves the sense of embodiment, but has no effect on the sense of presence in 360° VR movie experiences. The reason why there was no significant effect on the sense of presence by the hand gesture-based interaction might be due to the spatial presence felt by the user in a 360° VR movie. This should be further investigated in future research. The results also indicate that hand appearance partly improves the sense of embodiment. It only effects the sense of ownership and has no effect on the sense of agency. For the sense of presence, hand appearance also had no effect, which might be due to the quality of the real hands not being as good as what is seen with natural vision. The real hands were rendered using a point cloud, which even though elicit a stronger sense of ownership, the visual fidelity was not enough to create a sense of presence. The hand appearance results are similar to those of Lok et al. [LNWB03] who found that visual fidelity did not increase the sense of presence. However, similar to this study, their real hand representation was also not as good as in the real world. The hand segmentation we used still showed part of the real world scene surrounding the hands, which tended to break user presence. For the sense of agency, the results corroborated past research [LJ16], but were also in contrast with Argelaguet et al. [AHTL16]. The hand appearance most likely would not affect on the sense of agency, so future work on the sense of agency should be focused on the interaction aspects.

Limitations

While developing the prototype, there were a number of limitations faced which should be considered by future researchers. The real hands were visualised using a point cloud, which was discontinuous in places and showed visual artefacts of the real environment.

Rendering a 3D model of the hand from the captured data could remove these artefacts and result in a different sense of presence. The SoftKinetic camera has a relatively narrow field of view compared to the view on the HMD, which decreases immersion, so it would be interesting to explore the use of cameras with a wider field of view.

7. Conclusion and Future Work

This paper investigated the effects of adding hand gesture-based interaction to 360° movie watching experiences. A prototype system visualising the user's hands in an immersive 360° VR video was developed. The prototype system used a VR HMD with two sensors (Leap Motion and SoftKinetic RGB-D camera) that capture and track the user's hands to visualise them in the virtual environment and allow interacting with virtual objects. Using this system, an experiment was conducted to study the effects of hand gesture-based interaction and hand appearance on user presence and embodiment in a 360° VR movie. The results showed an increase in the sense of embodiment with interaction and real hand visualisation, however no effect on the sense of presence was discovered. The study also identified the limitations to the experiment.

For future research, several possibilities can be pursued. Among the four quadrants identified by Dolan and Parets [DP17], Participant Passive and Participant Active are the least explored. This study investigated the Participant Passive quadrant, so a similar study could be conducted for the Participant Active quadrant in which the user can engage physically with the story, changing the outcome of the narrative. This may lead to greater sense of presence and embodiment. As an example, carrying a birthday cake into a room of guests and being able to drop it on the way or present it to the birthday person. This would provide an option to change the narrative by physically interacting with it. Another possibility would be to compare the effects of offline rendered hand textures with the real-time rendered hand textures on the senses of presence and embodiment.

8. Acknowledgements

The authors would like to thank the Ministry of Business, Innovation and Employment, New Zealand for funding this research.

References

[AHTL16] Argelaguet F., Hoyet L., Trico M., Lecuyer A.: The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality (VR)* (March 2016), pp. 3–10. doi:10.1109/VR.2016.7504682. 2, 6, 7

[BGESB13] Bai H., Gao L., El-Sana J., Billinghurst M.: Free-hand interaction for handheld augmented reality using an rgb-depth camera. In *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications* (New York, NY, USA, 2013), SA '13, ACM, pp. 22:1–22:4. 3

[BM92] Besl P. J., McKay N. D.: A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell. 14*, 2 (Feb. 1992), 239–256. 3

[BSRH09] Bruder G., Steinicke F., Rothaus K., Hinrichs K.: Enhancing presence in head-mounted display environments by visual body feedback using head-mounted cameras. In *CyberWorlds, 2009. CW'09. International Conference on* (2009), IEEE, pp. 43–50. 2

[CWY*02] Cheok A. D., Weihua W., Yang X., Prince S., Wan F. S., Billinghurst M., Kato H.: Interactive theatre experience in embodied+ wearable mixed reality space. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality* (2002), IEEE Computer Society, p. 59. 2

[DP17] Dolan D., Parets M.: *Redefining The Axiom Of Story: The VR And 360 Video Complex*, (Accessed: August 21, 2017). https://techcrunch.com/2016/01/14/redefining-the-axiom-of-story-the-vr-and-360-video-complex/. 1, 8

[IKT*10] Ichikari R., Kikuchi K., Toishita W., Tenmoku R., Shibata F., Tamura H.: On-site real-time 3d match move for mr-based previsualization with relighting. In *ACM SIGGRAPH 2010 Talks* (New York, NY, USA, 2010), SIGGRAPH '10, ACM, pp. 36:1–36:1. doi:10.1145/1837026.1837074. 2

[Jae17] JaeYunLee: *Easy Movie Texture*, (Accessed: August 21, 2017). URL: https://www.assetstore.unity3d.com/en/#!/content/10032. 3, 4

[KEB*04] Krijn M., Emmelkamp P. M., Biemond R., de Wilde de Ligny C., Schuemie M. J., van der Mast C. A.: Treatment of acrophobia in virtual reality: The role of immersion and presence. *Behaviour Research and Therapy 42*, 2 (2004), 229 – 239. doi:10.1016/S0005-7967(03)00139-6. 6

[Lea17] LeapMotion: *Leap Motion VR Mount*, (Accessed: August 21, 2017). URL: https://www.thingiverse.com/thing:445866. 3

[LJ16] Lin L., Jörg S.: Need a hand?: How appearance affects the virtual hand illusion. In *Proceedings of the ACM Symposium on Applied Perception* (New York, NY, USA, 2016), SAP '16, ACM, pp. 69–76. doi:10.1145/2931002.2931006. 2, 7

[LNWB03] Lok B., Naik S., Whitton M., Brooks F.: Effects of handling real objects and avatar fidelity on cognitive task performance in virtual environments. In *Virtual Reality, 2003. Proceedings. IEEE* (2003), IEEE, pp. 125–132. 2, 7

[MBMSB15] McGill M., Boland D., Murray-Smith R., Brewster S.: A dose of reality: overcoming usability challenges in vr head-mounted displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), ACM, pp. 2143–2152. 2, 6

[Rea17] ReachVR S.: *DK2 BRACKET*, (Accessed: August 21, 2017). URL: http://www.reachvr.com/brackets. 3

[SFR01] Schubert T., Friedmann F., Regenbrecht H.: The experience of presence: Factor analytic insights. *Presence: Teleoperators and virtual environments 10*, 3 (2001), 266–281. 2, 6

[TIS*06] Tenmoku R., Ichikari R., Shibata F., Kimura A., Tamura H.: Design and prototype implementation of mr pre-visualization workflow. In *DVD-ROM Proc. Int. Workshop on Mixed Reality Technology for Filmmaking* (2006), pp. 1–7. 2

[UAW*99] Usoh M., Arthur K., Whitton M. C., Bastos R., Steed A., Slater M., Brooks Jr F. P.: Walking> walking-in-place> flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques* (1999), ACM Press/Addison-Wesley Publishing Co., pp. 359–364. 2

[UCAS00] Usoh M., Catena E., Arman S., Slater M.: Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments 9*, 5 (2000), 497–503. 2

[WDa17] WDallGraphics: *Lovebird Animated*, (Accessed: August 21, 2017). URL: https://www.assetstore.unity3d.com/en/#!/content/36951. 4

[WFGH11] Wobbrock J. O., Findlater L., Gergle D., Higgins J. J.: The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2011), CHI '11, ACM, pp. 143–146. doi:10.1145/1978942.1978963. 6